5.17

1E

STEP 1.

a) Prirate:

Benchmark A = Prirate miss Rate x Memory hit latency + (1 - Prirate miss rate) x prirate cache hit latency

$= 0.003 \times 120 + (1 - 0,003) \times 6$

$= 0,36 + 0,997 \times 6$

$= 0,36 + 5.982$

$= 6.342$

step 2

Benchmark B = Prirate miss rate x Memory hit latency + (1 - Private miss rate) x prirate cache hit latency.

$= 0.0006 \times 120 + (1 - 0,0006) \times 6$

$= 0,072 + 0,9994 \times 6$

$= 6,0684$

step 3

Benchmark A = shared miss rate x Memory hit latency + (1 - shared miss rate) x private cache hit latency

$= 0.0012 \times 120 + (1 - 0.0012) \times 12$

$= 0.144 + 0,9988 \times 12$

$= 12,1296$

step 4

Benchmark B = shared miss rate x Memory hit latency + (1 - shared miss Rate) x prirate cache hit latency.

$= 0,0012 \times 120 + (1 - 0,0012) \times 20$

$= 0,144 + 0,9988 \times 20$

$= 20,12$

step 5

b): Private

Benchmark A = Private missrate x Memory hit latency + (1 - Private miss Rate) x private cache hit latency.

= 0.003 x 120 + (1 - 0.003) x 6

= 0.36 + 0.997 x 6

= 0,36 + 5.982

= 6.342

step 6

Benchmark B = Private miss rate x Memory hit latency + (1 - Private miss Rate) x private cache hit latency.

= 0.0006 x 120 + (1 - 0,0006) x 6

= 0,072 + 0,9994 x 6

= 6,0684

step 7 shared:

Benchmark A = shared miss rate x Memory hit latency + (1 - Shad miss Rate) x private cache hit latency

= 0,0012 x 120 + (1 - 0,0012) x 12

= 0,144 + 0,9988 x 12

= 12.1296

step 8

Benchmark B = shared miss rate x memory hit latency + (1 - shared Miss Rate) x private cache hit latency.

= 0,0012 x 120 + (1 - 0,0012) x 20

= 0,144 + 0,9988 x 20

= 20,12

(2)

step 1

a) When shared L2 latency doubles, both benchmarks prefer private L2

step 2

(b) shared cache latency doubling:

Benchmark a/b: shared miss rate * memory hit latency + (1-shared miss rate) * 2 * shared cache hit latency.

Benchmarks A shared: 0,0012 * 120 + 0,9983 * 220 = 40.0096

Benchmarks B shared: 0,003 * 120 + 0,9997 * 2 * 220 = 40,024

Offchip memory latency doubling:

Benchmark a/b: private miss rate * 2 * memory hit latency + (1 - private miss rate) * private cache hit latency

Benchmark a/b: shared miss rate * 2 * memory hit latency + (1 - shared miss rate) * shared cache hit latency

5.17

(3)

a) Shared L2: typically good for multithreaded benchmarks when significant amount of shared data, good for applications need more than private cache capacity

Private L2: good for applications whose working set can fit, also good for isolating negative interfaces between multi programmed work loads.

(b) Shared L2: typically good for multithreaded benchmarks when significant amount of shared data good for applications need more than private cache capacity

Private L2

good for applications whose working set can fit, also good for isolating negative interfaces between multi programmed work loads.

6 5.17

Consider the given memory hierarchy, then divide 3 kinds of optimizations can improve the number of concurrent misses There are:

1) Processor: It is handle to multiple hardware threads, laarger load/store queue and out of order execution

2) Caches: It is more miss status handling registers.

3) Memory: Its support multiple outstanding memory requests by using memory controller.